CONFIDENCE INTERVALS FOR SOME FUNCTIONS OF SEVERAL BERNOULLI PARAMETERS WITH RELIABILITY APPLICATIONS[1]

by

Dar-Shong Hwang and Robert J. Buehler

Technical Report No. 161

November 1971

University of Minnesota
Minneapolis, Minnesota

[1]This research was supported by National Science Foundation Grant GP-24183.

CONFIDENCE INTERVALS FOR SOME FUNCTIONS OF SEVERAL BERNOULLI PARAMETERS WITH RELIABILITY APPLICATIONS[1]

By

Dar-Shong Hwang and Robert J. Buehler

University of Minnesota

[1]This research was supported by National Science Foundation Grant GP-24183.

ABSTRACT

In reliability theory one is often interested in the estimation of products and other functions of Bernoulli parameters. Previous work has been mainly concerned with binomial data. In the present paper other sampling methods are considered, negative binomial in particular. It is shown that the theory of exponential families can be used to obtain exact confidence limits for products and quotients of Bernoulli parameters when negative binomial observations are available from each population. For the case of two populations the relevant distributions are related to hypergeometric functions. To estimate more general functions such as sums of products of Bernoulli parameters, some sampling methods are suggested which are based on theory of compound distributions. Another method of obtaining exact confidence limits for products is given which exploits the independence of the minimum and difference of independent geometric variates. This last method is a discrete analog of a procedure due to Lieberman and Ross for estimating sums of scale parameters of exponential distributions.

ABSTRACT

In reliability theory one is often interested in the estimation of products and other functions of Bernoulli parameters. Previous work has been mainly concerned with binomial data. In the present paper other sampling methods are considered, negative binomial in particular. It is shown that the theory of exponential families can be used to obtain exact confidence limits for products and quotients of Bernoulli parameters when negative binomial observations are available from each population. For the case of two populations the relevant distributions are related to hypergeometric functions. To estimate more general functions such as sums of products of Bernoulli parameters, some sampling methods are suggested which are based on theory of compound distributions. Another method of obtaining exact confidence limits for products is given which exploits the independence of the minimum and difference of independent geometric variates. This last method is a discrete analog of a procedure due to Lieberman and Ross for estimating sums of scale parameters of exponential distributions.

1. Introduction.

Let $\pi_1, \pi_2, \ldots,$ denote Bernoulli populations with parameters $p_1, p_2, \ldots$. We are concerned here with interval estimation of various functions of the parameters, in particular, products, quotients, and sums of products. Most previous work in this area has been concerned with data in the form of binomial observations. The present paper shows that other sampling methods (negative binomial, for example) lead to new and possibly advantageous ways of obtaining confidence intervals. In some cases, sampling rules are suggested which depend upon the particular parametric function to be estimated.

1.1 Practical applications.

Our motivation comes partly from reliability theory and partly from biomedical applications. In reliability theory p_j may denote the reliability of a component from population π_j. Then a system made up of one component from each of k populations has reliability

$$R_1 = p_1 p_2 \cdots p_k \tag{1.1}$$

if the components are in series, or reliability

$$R_2 = 1 - q_1 q_2 \cdots q_k \qquad (q_i = 1 - p_i) \tag{1.2}$$

if they are in parallel. Other systems whose reliabilities are more general polynomials in the p_i are also of interest. In any case we wish to estimate the system reliability using data pertaining to the individual components.

Interval estimation of R_1 or R_2 (or special cases) has been considered by various writers. Buehler [2] and Harris [5] use Poisson approximations to estimate R_2. Madansky [12] and Myhre and Saunders

[14] estimate R_1 and R_2 by a chi-square approximation. Other work is reviewed by Rosenblatt [18] and by Mann [13].

Biomedical applications of the methods of Section 2 below have been considered by Hwang [6]. These include: (i) the estimation of the difference of two bacterial densities by the dilution method; (ii) comparison of two Yule's birth processes; and (iii) estimation of ratios and cross ratios of proportions, which arise, for example, in the theory of effectiveness indices. We hope to consider these in a later paper.

1.2 Nature of the difficulties.

In general we wish to find confidence limits for a given function of k Bernoulli parameters. Of course, large sample methods will always yield approximate solutions, but the present paper is concerned with mathematical devices for obtaining "exact" solutions despite the presence of $k - 1$ nuisance parameters.

The second difficulty arises from the discreteness of the distributions. For any confidence level γ, it is possible to give intervals with probability of coverage exactly equal to γ only by artificial randomization (see for example [19] or Section 3.5 of [8].). We prefer the usual "conservative" solutions having coverage probability greater than or equal to γ. But such solutions may justifiably be criticized if they are in a sense, too conservative. One quantitative measure of this would be the probability of coverage averaged over the parameter space. Hopefully this average should be only slightly greater than γ; but for discrete distributions having all (or nearly all) of the probability concentrated on only a few mass points, it may be closer to unity. It

[14] estimate R_1 and R_2 by a chi-square approximation. Other work is reviewed by Rosenblatt [10] and by Mann [1].

Biomedical applications of the methods of Section 2 below have been considered by Chang [6]. These include: (i) the estimation of the difference of two bacterial densities by the dilution method; (ii) comparison of two Yule's birth processes; and (iii) estimation of ratios and cross ratios of proportions, which arise, for example, in the theory of effectiveness indices. We hope to consider these in a later paper.

1.2. <u>Nature of the difficulties</u>.

In general we wish to find confidence limits for a given function of k Bernoulli parameters. Of course, large sample methods will always yield approximate solutions, but the present paper is concerned with mathematical devices for obtaining "exact" solutions despite the presence of $k - 1$ nuisance parameters.

The second difficulty arises from the discreteness of the distributions. For any confidence level $1 - \alpha$, it is possible to find intervals with probability of coverage exactly equal to $1 - \alpha$ only by artificial randomization (see for example [19] or Section 4.9 of [ ]). We refer the usual "conservative" solutions having coverage probability greater than or equal to $1 - \alpha$. But such solutions may justifiably be criticized if they are in a sense, too conservative. One quantitative measure of this would be the probability of coverage averaged over the parameter space. Hopefully this average should be only slightly greater than $1 - \alpha$; but for discrete distributions having all (or nearly all) of the probability concentrated on only a few mass points, it may be closer to unity. It will be convenient to have an expression for this undesirable state of affairs, and we will speak in a qualitative way of situations having a high "discreteness index" to indicate solutions which are highly "conservative." Examples would be a binomial distribution with small n ($n = 3$, say), or a Poisson distribution where prior information would indicate a mean small compared to unity.

A third difficulty is that in reliability applications in particular, the parameters p_j may be close to unity. This fact may have undesirable consequences for the discreteness index or for the expected sample size with certain sampling schemes.

It is probably impossible to deal simultaneously with all these difficulties in a completely satisfactory way, at least within the Neyman-Pearson framework. Nevertheless we hope that the methods given below will find a number of legitimate applications.

1.3 Summary.

In Section 2 we show that inverse (i.e., negative) binomial sampling allows us to use the Lehmann-Scheffé theory of exponential families to estimate products and quotients of Bernoulli parameters. Formulas are given for the most general case and for a number of special cases. Section 3 describes a new sampling method which is specifically aimed at the estimation of certain parametric functions not covered in Section 2. Section 4 gives a number of ways of using properties of compound Poisson distributions to estimate sums of products of Bernoulli parameters. In Section 5 we show that a method due to Lieberman and Ross for estimating sums of scale parameters of exponential distributions can be modified so that it applies also to problems involving products of Bernoulli parameters.

will be convenient to have an expression for this undesirable state of affairs, and we will speak in a qualitative way of situations having a high "discreteness index" to indicate solutions which are highly "conservative." Examples would be a binomial distribution with small n ($n = 1$, say), or a Poisson distribution where prior information would indicate a mean small compared to unity.

A third difficulty is that in reliability applications in particular, the parameters p_i may be close to unity. This fact may have undesirable consequences for the discreteness index or for the expected sample size with certain sampling schemes.

It is probably impossible to deal simultaneously with all these difficulties in a completely satisfactory way, at least within the Neyman-Pearson framework. Nevertheless we hope that the methods given below will find a number of legitimate applications.

1.4 Summary.

In Section 2 we show that inverse (i.e., negative) binomial sampling allows us to use the Lehmann-Scheffé theory of exponential families to estimate products and quotients of Bernoulli parameters. Formulas are given for the most general case and for a number of special cases. Section 3 describes a new sampling method which is specifically aimed at the estimation of certain parametric functions not covered in Section 2. Section 4 gives a number of ways of using properties of compound Poisson distributions to estimate sums of products of Bernoulli parameters. In Section 5 we show that a method due to Lieberman and Ross for estimating sums of scale parameters of exponential distributions can be modified so that it applies also to problems involving products of Bernoulli parameters.

2. Distribution Theory for Estimation of Products and Quotients Using Inverse Binomial Sampling.

In this section we exploit the theory of exponential families to remove nuisance parameters when the probability models are negative binomial. Other reliability applications of the same theory have been made by Lentner and Buehler [10] for gamma models and by Harris [5] for Poisson models.

2.1 The general case.

Let a Bernoulli population have probability p of success and for any fixed $r = 1, 2, \ldots,$ let X denote the number of successes observed before the r^{th} failure. Then X has the negative binomial distribution

$$(2.1) \qquad P\{X = x\} = \binom{r+x-1}{x} p^x (1-p)^r \qquad x = 0, 1, 2, \ldots$$

which we will abbreviate by

$$(2.2) \qquad X \sim NB(r, p).$$

If we sample independently from $k + k'$ Bernoulli populations to obtain random variables $X_i \sim NB(r_i, p_i)$ for $i = 1, \ldots, k$ and $Y_j \sim NB(s_j, p_j')$ for $j = 1, \ldots, k'$, then it will be seen to be possible to obtain confidence intervals for the function

$$(2.3) \qquad \theta = p_1 p_2 \cdots p_k / p_1' p_2' \cdots p_{k'}' \; .$$

The essential reasons are that the distribution (2.1) has the exponential form

$$(2.4) \qquad g(\varphi) h(x) e^{\varphi x}$$

where

2. Distribution Theory for Estimation of Products and Quotients Using Inverse Binomial Sampling.

In this section we exploit the theory of exponential families to remove nuisance parameters when the probability models are negative binomial. Other reliability applications of the same theory have been made by Lentner and Buehler [10] for gamma models and by Harris [ ] for Poisson models.

2.1 The general case.

Let a Bernoulli population have probability p of success and for any fixed $r = 1, 2, \ldots$, let X denote the number of successes observed before the r^{th} failure. Then X has the negative binomial distribution.

$$(2.1) \qquad P\{X = x\} = \binom{x+r-1}{x} p^r (1-p)^x \qquad x = 0, 1, 2, \ldots$$

which we will abbreviate by

$$(2.2) \qquad X \sim NB(r, p).$$

If we sample independently from $k + k'$ Bernoulli populations to obtain random variables $X_i \sim NB(r_i, p_i)$ for $i = 1, \ldots, k$ and $Y_j \sim NB(r_j', p_j')$ for $j = 1, \ldots, k'$, then it will be seen to be possible to obtain confidence intervals for the function

$$(2.3) \qquad \phi = p_1 p_2 \cdots p_k / p_1' p_2' \cdots p_{k'}' .$$

The essential reasons are that the distribution (2.1) has the exponential form

$$(2.4) \qquad \beta(\phi) h(x) e^{\theta x}$$

where

$$(2.5) \qquad \varphi = \log p, \ g(\varphi) = (1-e^{\varphi})^r, \ h(x) = \binom{r+x-1}{x}$$

and that $\log \theta$ is a linear function of $\varphi_i = \log p_i$ and $\varphi'_j = \log p'_j$. To apply the theory of Lehmann and Scheffé [9] (or see Sec. 4.4 of [8]), we transform the random variables by

$$(2.6) \qquad \begin{aligned} W_1 &= X_1 \\ W_i &= X_i - X_1 \qquad \text{for} \quad i = 2, 3, \dots, k, \\ V_j &= Y_j + X_1 \qquad \text{for} \quad j = 1, 2, \dots, k'. \end{aligned}$$

Let $U = (W_2, W_3, \dots, W_k, V_1, V_2, \dots, V_{k'})$ and $u = (w_2, w_3, \dots, w_k, v_1, v_2, \dots, v_{k'})$. Then it is straightforward to verify that

$$(2.7) \qquad P\{W_1 = w_1, U = u\} = A(u)B(w_1, u)\theta^{w_1}$$

where θ is defined by (2.3),

$$(2.8) \qquad A(u) = q_1^{r_1} \prod_{i=2}^{k} \prod_{j=1}^{k'} q_i^{r_i} p_i^{w_i} q_j'^{s_j} p_j'^{v_j}$$

$$(2.9) \qquad B(w_1, u) = \binom{r_1+w_1-1}{w_1} \prod_{i=2}^{k} \prod_{j=1}^{k'} \binom{r_i+w_i+w_1-1}{w_i+w_1}\binom{s_j+v_j-w_1-1}{v_j-w_1}$$

$$(2.10) \qquad q_i = 1 - p_i \ , \ q'_j = 1 - p'_j \ ,$$

and the range of the variables is

$$(2.11) \qquad \begin{aligned} w_1 &= 0, 1, 2, \dots \\ w_i &= -w_1, -w_1+1, -w_1+2, \dots \\ v_j &= w_1, w_1+1, w_1+2, \dots. \end{aligned}$$

From (2.7) we obtain the conditional distribution

$$(2.12) \qquad P\{W_1 = w_1 | U = u\} = B(w_1, u)\theta^{w_1} / \sum_t B(t, u)\theta^t$$

where the possible values of w_1 are $\max\{0, \max(-w_2, \dots, -w_k)\} \le w_1 \le \min(v_1, \dots, v_{k'})$, and the values of t in the summation are the same

as the possible values of w_1. Thus by using known theory of exponential families we have obtained a conditional distribution (2.12) which depends on the $k + k'$ parameters only through the function θ. Confidence intervals can be obtained from (2.12) in the usual way. For any confidence level γ, and any observation $W_1 = w_1$, $U = u$, an upper confidence limit $\theta_2(w_1, u)$ is defined by

$$(2.13) \qquad \theta_2(w_1, u) = \sup\{\theta: \sum_{t \le w_1} P\{W_1 = t | U = u\} \ge 1 - \gamma\}.$$

Similarly a lower confidence limit $\theta_1(w_1, u)$ is defined by

$$(2.14) \qquad \theta_1(w_1, u) = \inf\{\theta: \sum_{t \ge w_1} P\{W_1 = t | U = u\} \ge 1 - \gamma\}.$$

Because of the discreteness of the distribution these are "conservative" intervals satisfying the inequalities

$$(2.15) \qquad P\{\theta \le \theta_2(W_1, u) | U = u\} \ge \gamma \quad \text{for all} \quad \theta, u$$

$$(2.16) \qquad P\{\theta \ge \theta_1(W_1, u) | U = u\} \ge \gamma \quad \text{for all} \quad \theta, u.$$

If randomization is used to make the probabilities equal to γ, the resulting solutions are known to be "uniformly most accurate unbiased" (see Chapter 4 of [8]).

2.2 The product of k parameters.

In this section we give results which are appropriate for the special case when θ in (2.3) is replaced by

$$(2.17) \qquad \theta^* = p_1 p_2 \cdots p_k.$$

We then simply drop the variates Y_j and V_j and redefine U and u by $U = (W_2, W_3, \ldots, W_k)$, $u = (w_2, w_3, \ldots, w_k)$. The conditional distribution (2.12) then becomes as the possible values of w_1. Thus by using known theory of exponential families we have obtained a conditional distribution (2.12) which depends on the $k + k'$ parameters only through the function θ. Confidence intervals can be obtained from (2.12) in the usual way. For any confidence level γ, and any observation $W_1 = w_1$, $U = u$, an upper confidence limit $\bar{\theta}(w_1, u)$ is defined by

$$\bar{\theta}(w_1, u) = \sup\{\theta: \sum_{t \le w_1} P\{W_1 = t | U = u\} \ge 1 - \gamma\}. \tag{2.13}$$

Similarly a lower confidence limit $\underline{\theta}(w_1, u)$ is defined by

$$\underline{\theta}(w_1, u) = \inf\{\theta: \sum_{t \ge w_1} P\{W_1 = t | U = u\} \ge 1 - \gamma\}. \tag{2.14}$$

Because of the discreteness of the distribution these are "conservative" intervals satisfying the inequalities

$$P\{\theta \le \bar{\theta}(W_1, u) | U = u\} \ge \gamma \text{ for all } \theta, u \tag{2.15}$$

$$P\{\theta \ge \underline{\theta}(W_1, u) | U = u\} \ge \gamma \text{ for all } \theta, u. \tag{2.16}$$

If randomization is used to make the probabilities equal to γ, the resulting solutions are known to be "uniformly most accurate unbiased" (see Chapter 5 of [ ]).

2.3 The product of k parameters.

In this section we give results which are appropriate for the special case when θ in (2.3) is replaced by

$$\theta^* = p_1 p_2 \cdots p_k. \tag{2.17}$$

We then simply drop the variates Y_1 and V_1 and redefine U and u by $U = (W_2, W_3, \ldots, W_k)$, $u = (w_2, w_3, \ldots, w_k)$. The conditional distribution (2.12) then becomes

$$(2.18)\qquad P\{W_1 = w_1 | U = u\} = B^*(w_1, u)(\theta^*)^{w_1} \Big/ \sum_t B^*(t, u)(\theta^*)^t$$

where

$$(2.19)\qquad B^*(w_1, u) = \binom{r_1+w_1-1}{w_1} \prod_{i=2}^{k} \binom{r_i+w_i+w_1-1}{w_i+w_1}$$

and the range of w_1 (and t) is $\max\{0, -w_2, \ldots, -w_k\} \leq w_1 < \infty$. We now show some relationships between these expressions and hypergeometric functions.

2.3 The product of two parameters.

The usual definition of the hypergeometric function is [20]:

$$(2.20)\qquad F(a, b, c; z) = \{1 + \frac{a \cdot b}{1 \cdot c} z + \frac{a(a+1)b(b+1)}{1 \cdot 2 \cdot c(c+1)} z^2 + \frac{a(a+1)(a+2)b(b+1)(b+2)}{1 \cdot 2 \cdot 3 \cdot c(c+1)(c+2)} z^3 + \ldots\}.$$

Let us define

$$(2.21)\qquad f(a, b, c; z, n) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+n)\Gamma(b+n)}{\Gamma(c+n)} \frac{z^n}{n!},$$

where $\Gamma(x)$ is the gamma function. Then it is straightforward to verify that

$$(2.22)\qquad F(a, b, c; z) = \sum_{n=0}^{\infty} f(a, b, c; z, n).$$

If we further define

$$(2.23)\qquad F_x(a, b, c; z) = \sum_{n=0}^{x} f(a, b, c; z, n),$$

then $F_\infty(a, b, c; z) = F(a, b, c; z)$.

Going back to (2.18), and taking the special case $k = 2$, replacing θ^* by

$$(2.24)\qquad \lambda = p_1 p_2,$$

$$(2.18) \qquad P\{N_1^* = n_1, U = m\} = B^*(n_1, m)\binom{*}{n_1} \Big/ \sum_t B^*(t, m)\binom{*}{t}$$

where

$$(2.19) \qquad B^*(n_1, m) = \binom{n_1+m_1-1}{m_1} \prod_{i=2}^{k} \binom{n_1+m_1+\dots+m_i-1}{m_i}$$

and the range of n_1 (and t) is [illegible] $\le n_1 \le$ [illegible]. We now show some relationships between these expressions and hypergeometric functions.

2.3 The product of two parameters.

The usual definition of the hypergeometric function is [20]:

$$(2.20) \qquad F(a, b, c; z) = \{1 + \frac{a \cdot b}{1 \cdot c} z + \frac{a(a+1)b(b+1)}{1 \cdot 2 \cdot c(c+1)} z^2 + \frac{a(a+1)(a+2)b(b+1)(b+2)}{1 \cdot 2 \cdot 3 \cdot c(c+1)(c+2)} z^3 + \dots\}.$$

Let us define

$$(2.21) \qquad t(a, b, c; z; n) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+n)\Gamma(b+n)}{\Gamma(c+n)} \frac{z^n}{n!},$$

where $\Gamma(x)$ is the gamma function. Then it is straightforward to verify that

$$(2.22) \qquad F(a, b, c; z) = \sum_{n=0}^{\infty} t(a, b, c; z; n).$$

If we further define

$$(2.23) \qquad F_N(a, b, c; z) = \sum_{n=0}^{N} t(a, b, c; z; n),$$

then $F_\infty(a, b, c; z) = F(a, b, c; z)$.

Going back to (2.18), and taking the special case k = 2, replacing m^* by

$$(2.24) \qquad m = m_1 m_2$$

we find

$$(2.25)\qquad P\{X_1 = x_1 | X_2 - X_1 = w\} = B'(x_1, w)\lambda^{x_1} / \sum_t B'(t, w)\lambda^t$$

where

$$(2.26)\qquad B'(x_1, w) = \binom{r_1+x_1-1}{x_1}\binom{r_2+w+x_1-1}{w+x_1}$$

and where the range of x_1 (and t) is $\max(0, -w) \leq x_1 < \infty$.

It is now straightforward to verify that the conditional distribution can be written

$$(2.27)\qquad P\{X_1 \leq x | X_2 - X_1 = w\} = \begin{cases} F_x(r_1, r_2+w, 1+w; \lambda)/F(r_1, r_2+w, 1+w; \lambda) & \text{if } w \geq 0 \\ F_{x+w}(r_2, r_1-w, 1-w; \lambda)/F(r_2, r_1-w, 1-w; \lambda) & \text{if } w < 0 \end{cases}$$

where x is an integer ≥ 0 if $w \geq 0$; and an integer $\geq -w$ if $w < 0$. By analogy with the incomplete gamma and beta functions, we may call the distributions in (2.27) "incomplete hypergeometric functions."

If we define

$$(2.28)\qquad \alpha = \max(0, -w),\ \beta = \max(0, w)$$

then (2.27) can be put in the alternative form

$$(2.29)\qquad P\{X_1 \leq x | X_1 - X_2 = w\} = \frac{F_{x+\alpha}(r_1+\alpha,\ r_2+\beta,\ 1+|w|;\ \lambda)}{F(r_1+\alpha,\ r_2+\beta,\ 1+|w|;\ \lambda)}$$

$$x = \alpha, \alpha+1, \ldots,\quad w = 0, \pm 1, \pm 2, \ldots.$$

In reliability applications, $\lambda = p_1 p_2$ is the reliability of a series system, and the above formulas are relevant. To use the same theory to estimate the reliability $1 - q_1 q_2$ of a parallel system we must redefine X_1 and X_2 to be the number of failures prior to the r_1^{th} and r_2^{th} successes, and the relevant conditional distribution is

(2.30) $$P\{X_1 = x_1 | X_2 - X_1 = w\} = \frac{f(r_1+\alpha,\ r_2+\beta,\ 1+|w|;\ \mu,\ x_1-\alpha)}{F(r_1+\alpha,\ r_2+\beta,\ 1+|w|;\ \mu)}$$

$$x_1 = \alpha,\ \alpha+1, \ldots$$

$$w = 0,\ \pm 1,\ \pm 2, \ldots$$

where α and β are defined in (2.28) and where $\mu = q_1 q_2$. From (2.30) we can obtain confidence intervals for μ or for the reliability $1 - \mu$.

2.4 Some special cases.

Suppose $r_1 = r_2 = 1$, so that X_1 and X_2 are the numbers of successes before the first failure in each population. Then X_1 and X_2 have geometric distributions and either from (2.27) or by a more direct argument we find that whether $w \geq 0$ or $w < 0$,

(2.31) $$P\{\min(X_1, X_2) = x | X_2 - X_1 = w\} = (1-\lambda)\lambda^x \qquad x = 0, 1, 2, \ldots.$$

Since the last expression is the same for all w, we conclude that $\min(X_1, X_2)$ and $X_2 - X_1$ are independently distributed and that the unconditional distribution of $\min(X_1, X_2)$ is the geometric distribution given by (2.31). These facts have been noted previously by Ferguson [4].

In this case the upper and lower confidence limits given by (2.13) and (2.14) become

(2.32) $$\lambda_2(x) = \sup\{\lambda : \sum_{t=0}^{x} (1-\lambda)\lambda^t \geq 1 - \gamma\}$$

(2.33) $$\lambda_1(x) = \inf\{\lambda : \sum_{t=x}^{\infty} (1-\lambda)\lambda^t \leq 1 - \gamma\}$$

which can be solved explicitly to give

(2.34) $$\lambda_1 = (1-\gamma)^{1/x},\ \lambda_2 = \gamma^{1/(x+1)} .$$

The lower confidence limit λ_1 is of greater interest in reliability

$$\text{(2.30)}\qquad P\{X_1 = x_1 \mid X_2 - X_1 = w\} = \frac{F(x_1 + w, r_2 + \ldots \text{[illegible]})}{F(x_1 + r_1, r_2 + \ldots \text{[illegible]})}$$

$$x_1 = 0, 0+1, \ldots$$

$$x = 0, \pm 1, \pm 2, \ldots$$

where r and β are defined in (2.20) and where $\theta = \theta_1\theta_2$. From (2.?) we can obtain confidence intervals for θ or for the reliability $1 - \theta$.

2.5 Some special cases.

Suppose $r_1 = r_2 = 1$, so that X_1 and X_2 are the numbers of successes before the first failure in each population. Then X_1 and X_2 have geometric distributions and either from (2.27) or by a more direct argument we find that whether $w \geq 0$ or $w < 0$,

$$\text{(2.31)}\qquad P\{\min(X_1, X_2) = x \mid X_2 - X_1 = w\} = (1-\theta)\theta^x \qquad x = 0, 1, 2, \ldots$$

Since the last expression is the same for all w, we conclude that $\min(X_1, X_2)$ and $X_2 - X_1$ are independently distributed and that the unconditional distribution of $\min(X_1, X_2)$ is this geometric distribution given by (2.31). These facts have been noted previously by Ferguson [3].

In this case the upper and lower confidence limits given by (2.15) and (2.16) become

$$\text{(2.32)}\qquad \theta_2(x) = \sup\{\theta : \sum_{t=0}^{x} (1-\theta)\theta^t \geq 1 - \gamma\}$$

$$\text{(2.33)}\qquad \theta_1(x) = \inf\{\theta : \sum_{t=x}^{\infty} (1-\theta)\theta^t \geq 1 - \gamma\}$$

which can be solved explicitly to give

$$\text{(2.34)}\qquad \theta_1 = (1-\gamma)^{1/x},\; \theta_2 = \gamma^{1/(x+1)} .$$

The lower confidence limit θ_1 is of greater interest in reliability applications, and for large x, λ_1 is asymptotically $1 + (1/x)\log(1-\gamma)$. As an example if $x = 100$, $\alpha = 0.05$, then $\lambda_1 = 0.970$.

When $r_1 = r_2 = 1$, our general method of setting confidence limits amounts to using the unconditional distribution of $\min(X_1, X_2)$. In reliability applications where we have two dissimilar elements in series with system reliability $\lambda = p_1 p_2$, the use of $\min(X_1, X_2)$ is equivalent to testing individual series systems (rather than components), recording the first system failure, and using that observation to estimate $p_1 p_2$.

Next suppose r_1 is arbitrary but $r_2 = 1$. Then if $w = X_2 - X_1 \geq 0$,

$$P\{X_1 = x_1 | X_2 - X_1 = w\} = \binom{r_1+x_1-1}{x_1}(1-\lambda)^{r_1}\lambda^{x_1} \qquad x_1 = 0, 1, 2, \ldots, \tag{2.35}$$

which is a $NB(r_1, \lambda)$ distribution, and is incidentally free of w. If $w < 0$, then the conditional distribution is a truncated negative binomial,

$$P\{X_1 = x_1 | X_2 - X_1 = w\} = \frac{\binom{r_1+x_1-1}{x_1}(1-\lambda)^{r_1}\lambda^{x_1}}{\sum_{t=-w}^{\infty}\binom{r_1+x_1-1}{t}(1-\lambda)^{r_1}\lambda^{t}} \tag{2.36}$$

$$x_1 = -w_1, -w_1+1, -w_1+2, \ldots,$$

which is not free of w. If we define the incomplete beta function as usual by

$$I_x(m, n) = \frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)}\int_0^x t^{m-1}(1-t)^{n-1}dt \tag{2.37}$$

then the cumulative forms of (2.35) and (2.36) are respectively (see for example [7], [15]),

$$P\{X_1 \leq x | X_2 - X_1 = w\} = I_{1-\lambda}(r_1, x+1) \qquad x = 0, 1, 2, \ldots \tag{2.38}$$

and

applications, and for large n, λ_1 is asymptotically $1 + (1/n)\log(1-v)$.

As an example if $n = 100$, $v = 0.05$, then $\lambda_1 =$ [illegible].

When $r_1 = r_2 = 1$, our general method of setting confidence limits amounts to using the unconditional distribution of $\min(X_1, X_2)$. In reliability applications where we have two dissimilar elements in series with system reliability $\lambda = p_1 p_2$, the use of $\min(X_1, X_2)$ is equivalent to testing individual series systems (rather than components), recording the first system failure, and using that observation to estimate $p_1 p_2$.

Next suppose r_1 is arbitrary but $r_2 = 1$. Then if $w = X_2 - X_1 \geq 0$,

$$(2.35) \qquad P\{X_1 = x_1 \mid X_2 - X_1 = w\} = \binom{r_1+x_1-1}{x_1}(1-\lambda)^{r_1}\lambda^{x_1} \qquad x_1 = 0, 1, 2, \ldots$$

which is a $NB(r_1, \lambda)$ distribution, and is incidentally free of w.

If $w < 0$, then the conditional distribution is a truncated negative binomial,

$$(2.36) \qquad P\{X_1 = x_1 \mid X_2 - X_1 = w\} = \frac{\binom{r_1+x_1-1}{x_1}(1-\lambda)^{r_1}\lambda^{x_1}}{\sum_{x=-w}^{\infty}\binom{r_1+x-1}{x}(1-\lambda)^{r_1}\lambda^{x}}$$

$$x_1 = -w, -w+1, -w+2, \ldots$$

which is not free of w. If we define the incomplete beta function as usual by

$$(2.37) \qquad I_x(m, n) = \frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)}\int_0^x t^{m-1}(1-t)^{n-1}\,dt$$

then the cumulative forms of (2.35) and (2.36) are respectively (see for example [7], [15]),

$$(2.38) \qquad P\{X_1 \leq n \mid X_2 - X_1 = w\} = I_{1-\lambda}(r_1, n+1) \qquad n = 0, 1, 2, \ldots$$

and

$$P\{X_1 \leq x \mid X_2 - X_1 = w\} = \frac{I_{1-\lambda}(r_1, x+1) - I_{1-\lambda}(r_1, -w)}{1 - I_{1-\lambda}(r_1, -w)} \tag{2.39}$$

$$x = -w, -w+1, -w+2, \ldots.$$

Equation (2.38) and other formulas useful for evaluating the negative binomial distribution are considered in more detail in Section 5 below.

2.5 A numerical example.

Table 1 gives a numerical example of lower confidence limits, illustrating the use of the formulas of Section 2.3. The point estimates

Table 1

Example	r_1	r_2	x_1	x_2	95% lower confidence limit for p_1p_2	Point estimate of p_1p_2
A	5	3	45	57	0.7924	0.855
B	10	6	90	114	0.8068	0.855

for p_1p_2 were calculated from $x_1x_2/(r_1+x_1)(r_2+x_2)$. Since Example B is obtained by doubling the data values of Example A, the point estimates are the same. Example B has a confidence limit slightly closer to the point estimate, as one would expect. The confidence limits were found by iterating the calculation of the cumulative distribution (2.29) about 5 times, eventually determining the value of λ such that (2.29) takes the value 0.95. This was done on a CDC 6600 computer using a Fortran program and double precision (29 digits). Each iteration required evaluation of the infinite hypergeometric series (2.22). Individual terms were calculated recursively, and the series was arbitrarily declared to have converged as soon as an individual term a_n was less than 10^{-10}. Let

$$P\{X_1 \geq x \mid X_2 - X_1 = v\} = \frac{F_{N-1}(x_1, x+1) - F_{N-1}(x_1, -v)}{1 - F_{N-1}(x_1, -v)} \qquad (2.39)$$

$$x = -v, -v+1, -v+2, \ldots$$

Equation (2.39) and other formulas useful for evaluating the negative binomial distribution are considered in more detail in Section 4 below.

2.5 A numerical example.

Table 1 gives a numerical example of lower confidence limits, illustrating the use of the formulas of Section 2.4. The point estimates

Table 1

Example	X_1	X_2	N_1	N_2	95% lower confidence limit for $\rho_1\rho_2$	Point estimates of $\rho_1\rho_2$
A	5		45	57	.324	0.099
B	10		90	114	[illegible]	0.095

for $\rho_1\rho_2$ were calculated from $x_1 N_2/(x_1 + N_1)(x_2 + N_2)$. Since Example B is obtained by doubling the data values of Example A, the point estimates are the same. Example B has a confidence limit slightly closer to the point estimate, as one would expect. The confidence limits are found by iterating the calculation of the cumulative distribution (2.29) about 5 times, eventually determining the value of λ such that (2.29) takes the value 0.95. This was done on a CDC 6600 computer using a Fortran program and double precision (29 digits). Each iteration required evaluation of the infinite hypergeometric series (2.22). Individual terms were calculated recursively, and the series was arbitrarily declared to have converged as soon as an individual term a_n was less than 10^{-10}. The

$S_n = \sum_0^n a_j$ denote the cumulative sum. In Example A, with $\lambda = 0.7924$ the calculation stopped with $S_{201} = 16435.77$ and $a_{201} = 8.803 \times 10^{-11}$, and in Example B, with $\lambda = 0.8068$ it stopped with $S_{357} = 2320684000$ and $a_{357} = 8.948 \times 10^{-11}$. Computing time was approximately one second for each iteration.

3. A Method of Mixtures of Distributions for Estimating Other Functions.

In Section 2 we have shown how the Lehmann-Scheffé theory can be used to estimate products and quotients of Bernoulli parameters. Occasionally other functions may be of interest. For example, in reliability theory, when a system consists of combinations of series and parallel elements, then the reliability can be expressed as sums and differences of products of Bernoulli parameters. In the present section we will show that sometimes it is possible to tailor the sampling rule to the function to be estimated, and thereby eliminate nuisance parameters and obtain confidence limits.

Lemma 3.1.

Let $p_1 + q_1 = p_2 + q_2 = 1$, and let $\mathrm{Bin}(n, p)$ denote a binomial variate in the usual notation. If $X \sim \mathrm{NB}(r, q_1)$ and $(Y|X = n) \sim \mathrm{Bin}(n, p_2)$, then $Y \sim \mathrm{NB}(r, q_1 p_2 / (1-q_1 q_2))$.

Proof:

The probability generating function (PGF) of X is

$$\varphi_X(t) = \left(\frac{p_1}{1-q_1 t}\right)^r. \tag{3.1}$$

The PGF of $(Y|X = 1)$ is

$$\varphi_{Y|X=1}(t) = (q_2 + p_2 t). \tag{3.2}$$

By a theorem for the PGF of a random sum (Feller [3], p. 287), the PGF of Y is

$s_n = \sum^{n} a_i$ denote the cumulative sum. In Example 1, with $\lambda = 0.[illegible]$ the calculation stopped with $s_{21} = 10[illegible]3.77$ and $a_{21} = [illegible] \times 10^{-11}$, and in Example 2, with $\lambda = 4.000$ it stopped with $s_{37} = 272.[illegible]$ and $a_{37} = 9.948 \times 10^{-11}$. Computing time was approximately one second for each iteration.

3. A Method of Mixtures of Distributions for Estimating Other Functions.

In Section 2 we have shown how the [illegible] theory can be used to estimate products and quotients of Bernoulli parameters. Occasionally other functions may be of interest. For example, in reliability theory, when a system consists of combinations of series and parallel elements, then the reliability can be expressed as sums and differences of products of Bernoulli parameters. In the present section we will show that sometimes it is possible to tailor the sampling rule to the function to be estimated, and thereby eliminate nuisance parameters and obtain confidence limits.

Lemma 3.1.

Let $p_1 + q_1 = p_2 + q_2 = 1$, and let $\mathrm{Bin}(n, p)$ denote a binomial variate in the usual notation. If $X \sim \mathrm{NB}(r, q_1)$ and $(Y|X = n) \sim \mathrm{Bin}(n, p_2)$, then $Y \sim \mathrm{NB}(r, q_1 p_2 / (1-q_1 q_2))$.

Proof:

The probability generating function (PGF) of X is

$$g_X(s) = \left(\frac{p_1}{1-q_1 s}\right)^r. \tag{3.1}$$

The PGF of $(Y|X = 1)$ is

$$g_{Y|X=1}(s) = (q_2 + p_2 s). \tag{3.2}$$

By a theorem for the PGF of a random sum (Feller [4], p. 2[illegible]), the PGF of Y is

(3.3) $$\varphi_Y(t) = \varphi_X(\varphi_{Y|X=1}(t)) = \left(\frac{p_1}{1-q_1(q_2+p_2t)}\right)^r = \left(\frac{q}{1-pt}\right)^r$$

where

(3.4) $$q = \frac{p_1}{1-q_1q_2}\ ,\ p = 1 - q = \frac{q_1p_2}{1-q_1q_2}$$

which is the PGF of the distribution asserted in the lemma.

Since the marginal distribution of Y depends only on the parametric function p of (3.4), it is seen that we can make inferences about p by a sampling scheme which first takes a negative binomial sample from population π_1, then a (positive) binomial sample from π_2, where the size of the second sample depends on the outcome of the first sample.

We are unable to give an example where the parameters p or q would be of interest. However, by combining the present result with those of Section 2, it is possible to make inferences about a wide variety of parametric functions. For example, suppose we have a system with two elements in parallel connected to a third in series. The system reliability is

(3.5) $$\theta = (1-q_1q_2)p_3.$$

We can write

(3.6) $$\theta = q_1p_2p_3/p\ ,$$

and since this is of the form (2.3) (with appropriate relabeling of "successes" and "failures" in π_1) we can use the method of Section 2. In this case we will require the compound sampling described in Lemma 3.1 from populations π_1 and π_2 plus an independent negative binomial sample from each of the populations π_1, π_2, π_3.

sample from each of the populations π_1, π_2, π_3.

from populations π_1 and π_2 plus an independent negative binomial

In this case we will require the compound sampling described in Lemma 4.1

"successes" and "failures" in π_1) we can use the results of Section 2.

and since this is of the form (2.1) (with appropriate relabeling of

$$= p_1p_2^2/p \,, \qquad (4.[illegible])$$

we can write

$$= (1-q_1q_2)p_3 \,. \qquad (4.[illegible])$$

system reliability is

with two elements in parallel connected to a third in series. The

variety of parametric functions. For example, suppose we have a system

those of Section 2, it is possible to make inferences about a wide

would be of interest. However, by combining the present result with

We are unable to give an example where the parameters p or q

size of the second sample depends on the outcome of the first sample.

population π_1, then a (positive) binomial sample from π_2 where the

by a sampling scheme which first takes a negative binomial sample from

function p of (4.[illegible]), it is seen that we can make inferences about p

Since the marginal distribution of Y depends only on the parametric

which is the PGF of the distribution asserted in the lemma.

$$p = \frac{p_1}{1-q_1q_2} \,, \quad q = 1 - p = \frac{q_1p_2}{1-q_1q_2} \qquad (4.[illegible])$$

where

$$\phi_Y(t) = \phi_X(\phi_{Y|X=1}(t)) = \left(\frac{p_1}{1-q_1(q_2+p_2t)}\right)^r = \left(\frac{p}{1-qt}\right)^r \qquad (4.[illegible])$$

4. Compound Poisson Methods.

In the present section we give some new methods of eliminating nuisance parameters which apply to the estimation of many functions of Bernoulli parameters, in particular sums of products. Unfortunately the methods do have certain deficiencies, as we will indicate. Nevertheless we feel the techniques have mathematical interest and that possibly ways may eventually be found to circumvent or minimize the shortcomings. The following well-known result is the basis of the methods.

Lemma 4.1.

If $(X|Y = n) \sim \text{Bin}(n, p)$ and $Y \sim \text{Po}(\lambda)$(Poisson with mean λ), then $X \sim \text{Po}(\lambda p)$.

Thus we can "compound" binomial variates, converting them to (compound) Poisson variates. If we take a binomial sample whose size is determined randomly by a Poisson variate of known mean λ, the result is a Poisson variate having mean λp. The purpose of converting to Poisson is to allow later manipulations, but unfortunately the original Poisson observation introduces unwanted variability (see Section 4.2).

4.1 Horizontal and vertical compounding.

If we independently compound $\text{Bin}(n_i, p_i)$ variates by $\text{Po}(\lambda_i)$ variates we get independent $\text{Po}(\lambda_i p_i)$ variates. The Lehmann-Scheffé theory we have used in Section 2 applies also to Poisson distributions, and in particular it is possible to obtain conditional distributions depending only on the product of Poisson variates, in this case on $\Pi(\lambda_i p_i)$. Relevant formulas have been given by Harris [5]. Here $\Pi\lambda_i$ is known, so we can obtain confidence limits for Πp_i. We call the method of this paragraph "horizontal compounding."

4. Compound Poisson Methods.

In the present section we give some new methods of eliminating nuisance parameters which apply to the estimation of many functions of Bernoulli parameters, in particular sums of products. Unfortunately the methods do have certain deficiencies, as we will indicate. Nevertheless we feel the techniques have mathematical interest and that possibly ways may eventually be found to circumvent or minimize the shortcomings. The following well-known result is the basis of the methods.

Lemma 4.1.

If $(X|Y = n) \sim \mathrm{Bin}(n, p)$ and $Y \sim \mathrm{Po}(\lambda)$ (Poisson with mean λ), then $X \sim \mathrm{Po}(\lambda p)$.

Thus we can "compound" binomial variates, converting them to (compound) Poisson variates. If we take a binomial sample whose size is determined randomly by a Poisson variate of known mean λ, the result is a Poisson variate having mean λp. The purpose of converting to Poisson is to allow later manipulations, but unfortunately the original Poisson observation introduces unwanted variability (see Section 4.2).

4.1 Horizontal and vertical compounding.

If we independently compound $\mathrm{Bin}(n_i, p_i)$ variates by $\mathrm{Po}(\lambda_i)$ variates we obtain independent $\mathrm{Po}(\lambda_i p_i)$ variates. The Lehmann-Scheffé theory we have used in Section 2 applies also to Poisson distributions, and in particular it is possible to obtain conditional distributions, and in [illegible] on the product of Poisson variates, in this case on $\Pi(\lambda_i p_i)$. Relevant formulas have been given by Harris [5]. Here λ_i is known, so we can obtain confidence limits for Πp_i. We call the method of this paragraph "horizontal compounding."

Now suppose $X_2 \sim \text{Bin}(n_2, p_2)$, where n_2 is an observed value of $X_1 \sim \text{Bin}(n_1, p_1)$, and n_1 is an observed value of $\text{Po}(\lambda)$. Two applications of Lemma 4.1 give the marginal distribution $X_2 \sim \text{Po}(\lambda p_1 p_2)$. This "vertical compounding" can clearly be carried to any number of steps. Since λ is known, we need only tables of the Poisson distribution to find confidence limits for Πp_i.

To estimate sums of products of Bernoulli parameters one can appeal to the reproductive property of the Poisson distributions, and possibly also to the following lemma (Feller [3], p. 301, problem 3):

Lemma 4.2.

In Y Bernoulli trials, where $Y \sim \text{Po}(\lambda)$, the numbers of successes and failures are stochastically independent.

For example, θ in (3.5) can be rearranged to $\theta = p_3(p_1 + q_1 p_2)$. This suggests observing $Y = n$ from $\text{Po}(\lambda)$, $X_1 = x_1$ from $\text{Bin}(n, p_1)$, $X_2 = x_2$ from $\text{Bin}(n-x_1, p_2)$. Then $X_1 \sim \text{Po}(\lambda p_1)$ and $X_2 \sim \text{Po}(\lambda q_1 p_2)$ and X_1 and X_2 are independent. Thus $X_1 + X_2 \sim \text{Po}(\lambda p_1 + \lambda q_1 p_2)$, and taking X_3 from $\text{Bin}(x_1 + x_2, p_3)$ gives the desired unconditional distribution $X_3 \sim \text{Po}(\lambda p_3(p_1 + q_1 p_2))$.

It is clear that there are many ways to estimate sums of products by horizontal and vertical compounding. For expressions involving differences rather than sums, algebraic manipulations using the identities $p_i + q_i = 1$ can be used to reverse the signs. Hwang [6] has classified functions which can be estimated by these various techniques.

4.2 Critique.

In either horizontal or vertical compounding, the use of the variates Y having known Poisson distribution is rightly viewed with suspicion.

Now suppose $X_2 \sim Bin(n_1, p_2)$, where n_2 is an observed value of $X_1 \sim Bin(n_1, p_1)$, and n_1 is an observed value of $Po(\lambda)$. Two applications of Lemma 4.2 give the marginal distribution $X_2 \sim Po(\lambda p_1 p_2)$. This "vertical compounding" can clearly be carried to any number of steps. Since λ is known, we need only tables of the Poisson distribution to find confidence limits for λp_1.

To estimate sums of products of Bernoulli parameters one can appeal to the reproductive property of the Poisson distributions, and possibly also to the following lemma (Feller [3], p. [illegible], problem [illegible]):

Lemma 4.4.

In Y Bernoulli trials, where $Y \sim Po(\lambda)$, the numbers of successes and failures are stochastically independent.

For example, θ in (3.5) can be rewritten to $\theta = p_3(p_1 + q_1 p_2)$. This suggests observing $Y = n$ from $Po(\lambda)$, $X_1 = n_1$ from $Bin(n, p_1)$, $X_2 = n_2$ from $Bin(n-n_1, p_2)$. Then $X_1 \sim Po(\lambda p_1)$ and $X_2 \sim Po(\lambda q_1 p_2)$ and X_1 and X_2 are independent. Thus $X_1 + X_2 \sim Po(\lambda(p_1 + q_1 p_2))$, and taking X_3 from $Bin(n_1 + n_2, p_3)$ gives the desired unconditional distribution $X_3 \sim Po(\lambda p_3(p_1 + q_1 p_2))$.

It is clear that there are many ways to estimate sums of products by horizontal and vertical compounding. For expressions involving differences rather than sums, algebraic manipulations using the identities $p_i + q_i = 1$ can be used to reverse the signs. Hwang [?] has classified functions which can be estimated by these various techniques.

4.2 Critique.

In either horizontal or vertical compounding, the use of the variates Y having known Poisson distribution is rightly viewed with suspicion.

Technically, Y is an ancillary statistic since its distribution does not depend on the unknown parameters. Contrary to our suggested procedures, either Bayesian theory or the likelihood principle would demand inferences conditional on the observed value of Y.

The variates Y do indeed introduce undesirable variability, which is the price we pay for elimination of nuisance parameters. We can get some idea of the effect of the added variability by considering a simplified case. Take $Y \sim Po(\lambda)$ and $(X|Y = n) \sim Bin(n, p)$. We may compare the conditional (given $Y = n$) point estimator

$$(4.1) \qquad \hat{p} = X/n$$

with the unconditional estimator

$$(4.2) \qquad \tilde{p} = X/\lambda .$$

The simplest comparison is that of conditional and unconditional variances:

$$(4.3) \qquad Var(\hat{p}|Y = n) = pq/n \ ; \quad Var\ \tilde{p} = p/\lambda .$$

A measure of the efficiency of $\tilde{p}$ relative to $\hat{p}$ is the quotient $(pq/n)/(p/\lambda) = q\lambda/n$, in which the relevant factor is q, since λ/n is of the order of unity for large λ. We conclude that the additional variability introduced by Y is very serious when q is small, but not serious when p is small so that q is close to 1.

On the other hand when p is small, it becomes necessary to choose λ to compromise between large samples (λ large) and a high "discreteness index" (λp small; see Section 1.2).

In any compounding scheme it would be advisable to consider questions of efficiency and "discreteness index" at each stage.

Technically, Y is an ancillary statistic since its distribution does not depend on the unknown parameters. Contrary to our suggested principles, either Bayesian theory or the likelihood principal would demand inferences conditional on the observed value of Y.

The variates Y do indeed introduce undesirable variability which is the price we pay for elimination of nuisance parameters. We can get some idea of the effect of this added variability by considering a simplified case. Take $Y \sim P(\lambda)$ and $(X|Y = n) \sim \mathrm{Bin}(n, p)$. We may compare the conditional (given $Y = n$) point estimator

$$(4.1) \qquad \hat{p} = X/n$$

with the unconditional estimator

$$(4.2) \qquad \tilde{p} = X/\lambda.$$

The simplest comparison is that of conditional and unconditional variances:

$$(4.3) \qquad \mathrm{Var}(\hat{p}|Y = n) = pq/n \; ; \quad \mathrm{Var}\, \tilde{p} = p/\lambda .$$

A measure of the efficiency of $\tilde{p}$ relative to $\hat{p}$ is the quotient $(pq/n)/(p/\lambda) = q\lambda/n$, in which the relevant factor is q, since λ/n is of the order of unity for large λ. We conclude that the additional variability introduced by Y is very serious when q is small, but not serious when p is small so that q is close to 1.

On the other hand when p is small, it becomes necessary to choose a compromise between large samples (λ large) and a high "discreteness index" (λp small; see Section 1.2).

In any compounding scheme it would be advisable to consider question of efficiency and "discreteness index" at each stage.

5. Estimation of Products Using Geometric Variates.

A sequence of Bernoulli trials--that is, a Bernoulli process--gives a record of "successes" and "failures," which may conveniently be recorded as zeros and ones. Let $p = P(\text{success}) = P(0)$, $q = 1 - p = P(\text{failure}) = P(1)$. In a record such as 0111010001101 the information about p is traditionally summarized by "$X \sim \text{Bin}(13, p)$ and $X = 6$," if it was decided in advance to stop with the thirteenth observation, or by "$X \sim \text{NB}(7, p)$ and $X = 6$" if it was decided in advance to stop with the seventh failure. In the latter case we may alternatively report the observation of seven independent geometric variates: $(X_1, \ldots, X_7) = (1, 0, 0, 1, 3, 0, 1)$ where X_i is the number of successes between the $(i-1)^{st}$ and i^{th} failures. We write $X_i \sim \text{Geom}(p)$ where

(5.1) $$X \sim \text{Geom}(p) \text{ means } P(X = x) = (1-p)p^x \text{ for } x = 0, 1, 2, \ldots.$$

In the present section we exploit properties of geometric variates to estimate products of Bernoulli parameters. Our methods are discrete analogs of those proposed by Lieberman and Ross [11] for the estimation of sums of exponential parameters.

5.1 A method for estimating p_1p_2 and its theoretical basis.

The following easily verified lemma follows from Ferguson's characterization of the geometric distribution [4].

Lemma 5.1.

Let X and Y be independent, $X \sim \text{Geom}(p_1)$, $Y \sim \text{Geom}(p_2)$, and let $q_1 = 1 - p_1$, $q_2 = 1 - p_2$,

(5.2) $$U = \min(X, Y), \quad V = Y - X.$$

Then U and V are independent, $U \sim \text{Geom}(p_1p_2)$, and

3. Estimation of Products Using Geometric Variates.

A sequence of Bernoulli trials--that is, a Bernoulli process--gives a record of "successes" and "failures," which may conveniently be recorded as zeros and ones. Let $p = P(\text{success}) = P(0)$, $q = 1 - p = P(\text{failure}) = P(1)$. In a record such as 0111010001101 the information about p is traditionally summarized by "$N \sim \text{Bin}(13, p)$ and $N = 6$" if it was decided in advance to stop with the thirteenth observation; or by "$N \sim \text{NB}(7, p)$ and $N = 6$" if it was decided in advance to stop with the seventh failure. In the latter case we may alternatively report the observation of seven independent geometric variates: $(X_1, \ldots, X_7) = (1, 0, 0, 1, 3, 0, 1)$ where X_i is the number of successes between the $(i-1)$st and ith failures. We write $X_i \sim \text{Geom}(p)$ where

(3.1) $\quad X \sim \text{Geom}(p)$ means $P(X = x) = (1-p)p^x$ for $x = 0, 1, 2, \ldots$

In the present section we exploit properties of geometric variates to estimate products of Bernoulli parameters. Our methods are discrete analogs of those proposed by Bickel and Doksum [11] for the estimation of sums of exponential parameters.

3.1 A method for estimating p_1p_2 and its theoretical basis.

The following easily verified lemma follows from Ferguson's characterization of the geometric distribution [4].

Lemma 3.1.

Let X and Y be independent, $X \sim \text{Geom}(p_1)$, $Y \sim \text{Geom}(p_2)$, and let $q_1 = 1 - p_1$, $q_2 = 1 - p_2$.

(3.2) $\quad U = \min(X, Y), \quad V = X - Y.$

Then U and V are independent, $U \sim \text{Geom}(p_1p_2)$, and

$$(5.3)\qquad P\{V = v\} = \begin{cases} cp_2^{v} & v > 0 \\ cp_1^{-v} & v \leq 0 \end{cases} \qquad \text{where} \quad c = q_1q_2/(1-p_1p_2).$$

Since the distribution of U depends only on p_1p_2, one or more U-values could be used to estimate p_1p_2. However, there is "left over" information in the values of V which can be used for increased efficiency.

The variate V is rather like a $\mathrm{Geom}(p_2)$ variate when $V > 0$ and $-V$ is like $\mathrm{Geom}(p_1)$ when $V < 0$. The case $V = 0$ in fact presents a difficulty (which did not arise in the Lieberman-Ross continuous model) which we overcome by an asymmetric treatment, arbitrarily classifying $V = 0$ with the negative values. From (5.3) we have:

Lemma 5.2.

$(V-1|V > 0) \sim \mathrm{Geom}(p_2)$. $(-V|V \leq 0) \sim \mathrm{Geom}(p_1)$.

Now let us suppose $X_1,\ldots, X_m$ are independent $\mathrm{Geom}(p_1)$ variates and $Y_1,\ldots, Y_n$ are independent $\mathrm{Geom}(p_2)$ variates. Let $U_1 = \min(X_1, Y_1)$, $V_1 = Y_1 - X_1$. We now use V_1 to construct an observation U_2, independent of U_1 and having the same $\mathrm{Geom}(p_1p_2)$ distribution.

Let us define

$$(5.4)\qquad U_2 = \begin{cases} \min(X_2, V_1-1) & \text{if } V_1 > 0 \\ \min(-V_1, Y_2) & \text{if } V_1 \leq 0 \end{cases}$$

$$(5.5)\qquad V_2 = \begin{cases} V_1 - 1 - X_2 & \text{if } V_1 > 0 \\ Y_2 + V_1 & \text{if } V_1 \leq 0\ . \end{cases}$$

From Lemmas 5.1 and 5.2 we have:

Lemma 5.3.

(i) U_1, U_2 and V_2 are mutually independent; (ii) $U_2 \sim \mathrm{Geom}(p_1p_2)$; (iii) V_2 has the distribution (5.3).

$$(5.3) \qquad P\{V = v\} = \begin{cases} c p_1^{-v} & v \le 0 \\ c p_2^{v} & v > 0 \end{cases} \qquad \text{where} \quad c = q_1 q_2/(1-p_1 p_2).$$

Since the distribution of U depends only on p_1p_2, one or more U-values could be used to estimate p_1p_2. However, there is "left over" information in the values of V which can be used for increased efficiency.

The variate V is rather like a $\text{Geom}(p_2)$ variate when $V > 0$ and $-V$ is like $\text{Geom}(p_1)$ when $V < 0$. The case $V = 0$ in fact presents a difficulty (which did not arise in the Lindeman-Ross continuous model) which we overcome by an asymmetric treatment, arbitrarily classifying $V = 0$ with the negative values. From (5.3) we have:

Lemma 5.2.

$(V-1|V > 0) \sim \text{Geom}(p_2)$, $(-V|V \le 0) \sim \text{Geom}(p_1)$.

Now let us suppose $X_1, \ldots, X_n$ are independent $\text{Geom}(p_1)$ variates and $Y_1, \ldots, Y_n$ are independent $\text{Geom}(p_2)$ variates. Let $U_1 = \min(X_1, Y_1)$, $V_1 = Y_1 - X_1$. We now use V_1 to construct an observation U_2, independent of U_1 and having the same $\text{Geom}(p_1p_2)$ distribution.

Let us define

$$(5.4) \qquad U_2 = \begin{cases} \min(X_2, V_1-1) & \text{if } V_1 > 0 \\ \min(-V_1, Y_2) & \text{if } V_1 \le 0 \end{cases}$$

$$(5.5) \qquad V_2 = \begin{cases} V_1 - 1 - X_2 & \text{if } V_1 > 0 \\ Y_2 + V_1 & \text{if } V_1 \le 0. \end{cases}$$

From Lemmas 5.1 and 5.2 we have:

Lemma 5.3.

(i) U_1, U_2 and V_2 are mutually independent; (ii) $U_2 \sim \text{Geom}(p_1p_2)$;

(iii) V_2 has the distribution (5.3).

At this point we have two observations U_1, U_2, useful for estimating p_1p_2, plus independent "left over" information V_2 which, like V_1, can be combined with a $\text{Geom}(p_1)$ observation if $V_2 > 0$ or with a $\text{Geom}(p_2)$ observation if $V_2 \leq 0$. The procedure continues until either the X's or the Y's are exhausted. As in the Lieberman-Ross method there will then be residual information about one parameter, but not about both. Suppose the procedure yields r values U_j. These are combined without loss of information about p_1p_2 by forming the sum

$$S_r = \sum_{i=1}^{r} U_i \ . \tag{5.6}$$

Well known distribution theory gives $S_r \sim NB(r, p_1p_2)$, so that it becomes a straightforward problem to find confidence limits for p_1p_2 using the negative binomial distribution, as we show in Section 5.4.

5.2 A symmetric approximation.

Some simplification is possible, and the results are more closely analogous to those of Lieberman and Ross, if we define the "symmetric approximation" to be that obtained by deleting the "-1" in the definitions (5.4) and (5.5) of U_2 and V_2. When q_1 and q_2 are small, then $V = 0$ has small probability, and the "symmetric approximation" is presumably appropriate. Let us define the following partial sums:

$$\begin{aligned} T_{1j} &= \sum_{i=1}^{j} X_i \qquad j = 1,\ldots, m \\ T_{2j} &= \sum_{i=1}^{j} Y_i \qquad j = 1,\ldots, n \\ S_j &= \sum_{i=1}^{j} U_i \qquad j = 1,\ldots, r' \end{aligned} \tag{5.7}$$

At this point we have two observations U_1, U_2, useful for estimating $p_1 p_2$, plus independent "left over" information V_2 which, like V_1, can be combined with a Geom(p_1) observation if $V_2 > 0$ or with a Geo(p_2) observation if $V_2 \leq 0$. The procedure continues until either the Y's or the Y's are exhausted. As in the Lieberman-Ross method there will then be residual information about one parameter, but not about both.

Suppose the procedure yields r values U_i. These are combined without loss of information about $p_1 p_2$ by forming the sum

$$S_r = \sum_{i=1}^{r} U_i . \qquad ([illegible])$$

All known distribution theory gives $S_r \sim NB(r, p_1 p_2)$, so that it becomes a straightforward problem to find confidence limits for $p_1 p_2$ using the negative binomial distribution, as we show in Section [illegible].

[illegible].2 A symmetric approximation.

Some simplification is possible and the results are more closely analogous to those of Lieberman and Ross, if we define the "symmetric approximation" to be that obtained by deleting the "-1" in the definition ([illegible]) and (5.5) of U_i and V_2. When p_1 and p_2 are small, then $V = 0$ has small probability and the "symmetric approximation" is presumably appropriate. Let us define the following partial sums:

$$T_{1j} = \sum_{i=1}^{j} X_i \qquad j = 1, \ldots, n$$

$$T_{2j} = \sum_{i=1}^{j} Y_i \qquad j = 1, \ldots, m \qquad (5.[illegible])$$

$$S_j = \sum_{i=1}^{j} U_i \qquad j = 1, \ldots, r.$$

where r' is the number of U-values provided by the symmetric approximation. As in the Lieberman-Ross procedure, it can be shown that the construction is such that the sequence of S_j's agrees exactly with the first r' terms obtained by putting the T_{1j}'s and T_{2j}'s into a single ordered sequence (an example is given in Section 5.3). Therefore $W_{r'}$ can be calculated simply by

$$(5.8) \qquad S_{r'} = \min(T_{1m}, T_{2n})$$

and r' is the number of T_{1j}'s and T_{2j}'s less than or equal to $S_{r'}$.

This approximation shows that when q_1 and q_2 are small there will be a minimum of "left over" information, so that the procedure has a high efficiency, when we have approximately the same number of Bernoulli observations from the two populations. Of course, it is also possible to deliberately terminate sampling in such a way as to minimize "left over" information.

Because of the simplifications arising in the symmetric approximation, it is tempting to think that it may actually be exact. A simple check shows that unfortunately it is not. Let $A = \{X_1 = Y_1 = 0\}$, $B = \{X_1 = X_2 = 0\}$, $C = \{Y_1 = Y_2 = 0\}$. Then with the modified definition of U_2, $\{U_1 = U_2 = 0\} = A \cup B \cup C$. Since $BC = ABC$, $P\{A \cup B \cup C\} = P(A) + P(B) + P(C) - P(AB) - P(AC) = q_1q_2 + q_1^2 + q_2^2 - q_1^2q_2 - q_1q_2^2$. When $q_1 = q_2 = q = 1 - p$, this gives $P\{U_1 = U_2 = 0\} = q^2(1+2p)$. But if U_1 and U_2 are independent $\mathrm{Geom}(p^2)$ variates, $P\{U_1 = U_2 = 0\} = (1-p^2)^2 = q^2(1+2p+p^2)$, so that the symmetric approximation is not exact.

5.3 <u>Examples</u>.

To illustrate a case where p_1 and p_2 are close to unity, let $(X_1, \ldots, X_3) = (24, 35, 39)$, $(Y_1, \ldots, Y_6) = (16, 12, 27, 12, 43, 19)$.

where r' is the number of U-values provided by the symmetric approximation. As in the Liebermann-Ross procedure, it can be shown that the construction is such that the sequence of δ_j's agrees exactly with the first r' terms obtained by putting the T_{1i}'s and T_{2i}'s into a single ordered sequence (an example is given in Section 5.4). Therefore $W_{r'}$ can be calculated simply by

$$(5.8) \qquad W_{r'} = \min(T_{1n'}, T_{2n''})$$

and r' is the number of T_1's and T_2's less than or equal to $\delta_{r'}$.

This approximation shows that when p_1 and p_2 are small there will be a minimum of "leftover" information, so that the procedure has a high efficiency, when we have approximately the same number of Bernoulli observations from the two populations. Of course, it is also possible to deliberately terminate sampling in such a way as to minimize "left over" information.

Because of the simplifications arising in the symmetric approximation, it is tempting to think that it may actually be exact. A simple check shows that unfortunately it is not. Let $A = \{Y_1 = X_1 = 0\}$, $C = \{X_1 = X_2 = 0\}$, $B = \{Y_1 = Y_2 = 0\}$. Then with the modified definition of U_2, $\{U_1 = U_2 = 0\} = A \cup B \cup C$. Since $BC = ABC$, $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) = q_1 q_2 + q_1^2 + q_2^2 - q_1^2 q_2 - q_1 q_2^2$. When $p_1 = p_2 = p = 1 - q$, this gives $P\{U_1 = U_2 = 0\} = q^2(1+2q)$. But if U_1 and U_2 are independent $\mathrm{Geom}(p^2)$ variates, $P\{U_1 = U_2 = 0\} = (1-p^2)^2 = q^2(1+q)^2$, so that the symmetric approximation is not exact.

5.4 Examples.

To illustrate a case where p_1 and p_2 are close to unity, let $(T_1, \ldots, T) = (24, 15, 39)$, $(T_1, \ldots, T) = (26, 12, 27, 17, 4, 19)$.

$U_1 = \min(24, 16) = 16$, $V_1 = 16 - 24 = -8$, and we take $+8$ to be a $\text{Geom}(p_1)$ observation. $U_2 = \min(8, 12) = 8$, $V_2 = 12 - 8 = 4$, and we take $4 - 1 = 3$ to be a $\text{Geom}(p_2)$ observation. Similarly we find: $(V_1,\ldots, V_7) = (-8, 4, -32, -5, 7, -33, 10)$, $(U_1,\ldots, U_7) = (16, 8, 3, 27, 5, 6, 33)$, $(S_1,\ldots, S_7) = (16, 24, 27, 54, 59, 65, 98)$. The "left over" information is $V_7 = 10$ and $Y_6 = 19$. The symmetric approximation would give $(S_1,\ldots, S_7) = (16, 24, 28, 55, 59, 67, 98)$, where $S_1 = Y_1$, $S_2 = X_1$, $S_3 = Y_1 + Y_2$, $S_4 = Y_1 + Y_2 + Y_3$, etc. We know $S_7 \sim NB(7, p_1p_2)$, and the value $S_7 = 98$ summarizes the information about p_1p_2. In effect, we have terminated with the seventh failure on the 105-th Bernoulli trial. The maximum likelihood estimator of p_1p_2 based on S_7 is 98/105, and (as we show in Section 5.4) a 90 percent confidence interval for p_1p_2 is (0.890,0.968).

With p_1 and p_2 small, we might get observations like $(X_1,\ldots, X_7) = (0, 1, 1, 0, 0, 2, 1)$ and $(Y_1,\ldots, Y_{12}) = (0, 0, 1, 0, 1, 1, 0, 2, 0, 0, 1, 2)$. Our asymmetrical solution gives $U_6 = U_{10} = U_{17} = 1$ and all other U-values equal to zero. Left over information is $X_7 = 1$ and $V_{17} = -1$. The maximum likelihood estimator of p_1p_2 based on $S_{17} = 3$ is $3/20 = 0.15$. The symmetric approximation gives a substantially different sequence of U values, and in fact direct comparison of the methods is difficult because the asymmetric solution tends to use a greater proportion of Y values, so that at termination the two methods have used different sets of data. We believe the symmetric approximation will tend to overestimate p_1p_2 when p_1 and p_2 are small.

In the above sequence $(X_1,\ldots, X_6)$ we have 4 successes and 6 failures, giving $\hat{p}_1 = 0.4$. In $(Y_1,\ldots, Y_{12})$ we have 8 successes and

$U_1 = \min(24, 15) = 15$, $V_1 = 15 - 24 = -9$, and we take +4 to be a Geo(p_1) observation. $U_2 = \min(8, 12) = 8$, $V_2 = 12 - 8 = 4$, and we take $4 - 1 = 3$ to be a Geo(p_2) observation. Similarly we find: $(V_1, \ldots, V_7) = (-9, 4, -22, -5, [illegible], [illegible], 10)$, $(U_1, \ldots, U_7) = (15, 8, [illegible], 27, 5, 6, 35)$, $(S_1, \ldots, S_7) = (15, 24, 27, 54, 59, 65, 90)$. The "left over" information is $V_7 = 10$ and $Y_6 = 1$. The asymptotic approximation would give $(S_1, \ldots, S_7) = (15, 24, 26, [illegible], 59, 67, 96)$, where $S_1 = Y_1$, $S_2 = Y_1 + Y_2$, $S_3 = Y_1 + Y_2 + Y_3$, etc. We know $S_i \sim NB(i, p_1p_2)$, and the value $S_7 = 90$ summarizes the information about p_1p_2. In effect, we have terminated with the seventh failure on the 10[illegible]th Bernoulli trial. The maximum likelihood estimator of p_1p_2 based on S_7 is 90/105, and (as we saw in Section 3.4) a [illegible] percent confidence interval for p_1p_2 is ([illegible]).

With p_1 and p_2 small, we might get observations like $(X_1, \ldots, X_7) = (0, 1, 4, 0, 0, 2, 1)$ and $(Y_1, \ldots, Y_{12}) = (0, [illegible], 1, 1, 0, 2, 1, 0, 2, 0, 0, 1, 2)$. Our asymmetrical solution gives $U_3 = U_{10} = U_{11} = 1$ and all other U-values equal to zero. Left over information is $X_7 = 1$ and $V_{11} = -1$. The maximum likelihood estimator of p_1p_2 based on S_{11} = [illegible] is 3/20 = 0.15. The asymmetric approximation gives a substantially different sequence of U values, and in fact direct comparison of the methods is difficult because the asymmetric solution tends to use a greater proportion of Y values, so that at termination the two methods have used different sets of data. We believe the asymmetric approximation will tend to overestimate p_1p_2 when p_1 and p_2 are small.

In the above sequence $(X_1, \ldots, X_7)$ we have [illegible] successes and [illegible] failures, giving $\hat{p}_1 = 0.4$. In $(Y_1, \ldots, Y_{12})$ we have 8 successes and

12 failures, giving $\hat{p}_2 = 0.4$. The value $\hat{p}_1\hat{p}_2 = 0.16$ agrees well with the value 0.15 derived above. We give this comparison simply as a partial check, and not to claim any superior method of point estimation. The point of the procedure is rather the possibility of finding "exact" confidence limits for p_1p_2. In the next section we find a 90 percent confidence interval (0.044, 0.344).

5.4 Calculation of negative binomial confidence limits.

Let us denote by $F_X(x; r, p)$ the cumulative distribution function of $X \sim NB(r, p)$ defined in (2.1). Then using standard arguments

$$(5.8) \qquad P\{p \leq \bar{p}(X)\} \geq \gamma \quad \text{and} \quad P\{\underline{p}(X) \leq p\} \geq \gamma$$

when $\bar{p}(x)$ and $\underline{p}(x)$ are solutions of

$$(5.9) \qquad F_X(x; r, \bar{p}) = 1 - \gamma \quad \text{and} \quad F_X(x-1; r, \underline{p}) = \gamma .$$

It is known [7], [15] that the negative binomial is related to the incomplete beta function defined in (2.37) by

$$(5.10) \qquad F_X(x; r, p) = I_{1-p}(r, x+1).$$

Therefore we may also define $\bar{p}$ and $\underline{p}$ as solutions of

$$(5.11) \qquad I_{1-\bar{p}}(r, x+1) = 1 - \gamma \quad \text{and} \quad I_{1-\underline{p}}(r, x) = \gamma .$$

Thus the confidence limits could be found from tables of the incomplete beta function [17]. It is usually more convenient however to use the F distribution. The relationship here (see for example [16], p. 33) is

$$(5.12) \qquad P\{F_{2n}^{2m} \geq F_0\} = I_c(n, m), \quad c = m/(m+nF_0),$$

where of course F_{2n}^{2m} is the Fisher-Snedecor F variate with 2m and 2n degrees of freedom. Thus we find

12 failures, giving $\hat{p}_2 = 0.4$. The value $\hat{p}_1\hat{p}_2 = 0.16$ agrees well with the value 0.15 derived above. We give this comparison simply as a partial check, and not to claim any superior method of point estimation. The point of the procedure is rather the possibility of finding "exact" confidence limits for p_1p_2. In the next section we find a 90 percent confidence interval (0.044, 0.44).

5.4 Calculation of negative binomial confidence limits.

Let us denote by $F_N(x; r, p)$ the cumulative distribution function of $X \sim N(r, p)$ defined in (2.1). Then using standard arguments

$$P\{p \le \bar{p}(X)\} \ge \gamma \text{ and } P\{\underline{p}(X) \le p\} \ge \gamma \tag{5.8}$$

when $\bar{p}(x)$ and $\underline{p}(x)$ are solutions of

$$F_N(x; r, \bar{p}) = 1 - \gamma \text{ and } F_N(x-1; r, \underline{p}) = \gamma. \tag{5.9}$$

It is known [7], [15] that the negative binomial is related to the incomplete beta function defined in (2.17) by

$$F_N(x; r, p) = I_{1-p}(r, x+1). \tag{5.10}$$

Therefore we may also define $\bar{p}$ and $\underline{p}$ as solutions of

$$I_{1-\bar{p}}(r, x+1) = 1 - \gamma \text{ and } I_{1-\underline{p}}(r, x) = \gamma. \tag{5.11}$$

Thus the confidence limits could be found from tables of the incomplete beta function [17]. It is usually more convenient however to use the F distribution. The relationship here (see for example [1], p. 74) is

$$P\{F^{2x}_{2r} \le F_0\} = I_c(r, x), \quad c = r/(r+xF_0), \tag{5.12}$$

where of course F^{2x}_{2r} is the Fisher-Snedecor F variate with 2x and 2r degrees of freedom. Thus we find

$$(5.13)\qquad \bar{p} = \frac{(x+1)F_{2r}^{2x+2}(\gamma)}{r + (x+1)F_{2r}^{2x+2}(\gamma)}\ , \quad \underline{p} = \frac{xF_{2r}^{2x}(1-\gamma)}{r + xF_{2r}^{2x}(1-\gamma)}$$

where $F_n^m(\beta)$ is the F-percentile defined by

$$P\{F_n^m \leq F_n^m(\beta)\} = \beta .$$

Since F tables usually give upper tails, we may use the standard trick of taking reciprocals to express $\underline{p}$ as

$$(5.14)\qquad \underline{p} = \frac{x}{x + rF_{2x}^{2r}(\gamma)} .$$

If neither the incomplete beta nor the F table is convenient, a number of approximations are available [1], [7].

For the first example of Section 5.3, $r = 7$, $x = 98$, $F_{14}^{198}(0.95) = 2.16$, $F_{196}^{14}(0.95) = 1.74$, and we find a 90 percent confidence interval $(\underline{p}, \bar{p}) = (0.890, 0.968)$. For the second example $r = 17$, $x = 3$, $F_{34}^{8}(0.95) = 2.23$, $F_{6}^{34}(0.95) = 3.79$, and $(\underline{p}, \bar{p}) = (0.044, 0.344)$ is a 90 percent confidence interval.

5.5 Estimation of $p_1p_2 \cdots p_k$.

There are many ways to extend the above procedures in order to estimate the product of k parameters. If we consider the symmetric approximation and let T_{in_i} denote the cumulative sum of all n_i geometric observations from population i, then $S_{r'} = \min_i(T_{in_i})$ is approximately $NB(r', p_1p_2 \cdots p_k)$ where r' is the number of cumulative sums $T_{ij}(j = 1, \ldots, n_i)$ less than or equal to $S_{r'}$. This analog of an exact result of Lieberman and Ross would appear to be a somewhat questionable approximation for large k.

Perhaps the simplest exact procedure is a (k-1)-fold iteration of our two-population procedure. Let $(X_{i1}, \ldots, X_{in_i})$ denote $Geom(p_i)$

$$\overline{p} = \frac{(x+1)F^{2x+2}_{2N}(\gamma)}{x + (x+1)F^{2x+2}_{2N}(\gamma)}\ , \quad \underline{p} = \frac{xF^{2N}_{2x}(1-\gamma)}{x + xF^{2N}_{2x}(1-\gamma)} \tag{5.12}$$

where $F^{m}_{n}(\delta)$ is the δ-percentile defined by

$$P\{F^{m}_{n} \leq F^{m}_{n}(\delta)\} = \delta .$$

Since F tables usually give upper tails, we may use the standard trick of taking reciprocals to express $\underline{p}$ as

$$\underline{p} = \frac{x}{x + NF^{2x}_{2N}(\gamma)} . \tag{5.13}$$

If neither the incomplete beta nor the F tables is convenient, a number of approximations are available [1], [7].

For the first example of Section 5.[illegible]: $N = 7$, $x = [illegible]$, $F^{[illegible]}_{14}(0.95) = 2.[illegible]$, $F^{14}_{10}(.[illegible]) = 1.7[illegible]$, and we find a [illegible] percent confidence interval $(\underline{p}, \overline{p}) = (0.[illegible], 0.[illegible])$. For the second example $N = 17$, $x = [illegible]$, $F^{[illegible]}_{[illegible]}(0.95) = 2.2[illegible]$, $F^{[illegible]}_{[illegible]}(0.95) = [illegible]$, and $(\underline{p}, \overline{p}) = (.[illegible], 0.[illegible])$ is a [illegible] percent confidence interval.

5.5 Estimation of $p_1p_2 \ldots p_k$.

There are many ways to extend the above procedures in order to estimate the product of k parameters. If we consider the Bayesian approximation and let T_{n_i} denote the cumulative sum of all n_i geometric observations from population i, then $\hat{q}_i = \min_i(T_{n_i})$ is approximately $NB(r; p_1p_2 \ldots p_k)$ where r is the number of cumulative sums T_{ij} ($j = 1, \ldots, n_i$) less than or equal to $\hat{q}_i$. This analog of an exact result of Lieberman and Ross would appear to be a somewhat questionable approximation for large k.

Perhaps the simplest exact procedure is a $(k-1)$-fold iteration of our two-population procedure. Let $(X_{i1}, \ldots, X_{in_i})$ denote Geom(p_i)

observations. Utilizing observations with $i = 1, 2$, we may construct by the method of Section 5.1, values $U_1, \ldots, U_r$ which are independent $\text{Geom}(p_1 p_2)$ variates. Next we combine $U_1, \ldots, U_r$ with $X_{31}, \ldots, X_{3n_3}$ to obtain $\text{Geom}(p_1 p_2 p_3)$ variates, and so on.

Other procedures can be described which involve choosing the minimum of more than two geometric variates. It does not seem worthwhile to spell out details here, but we will state without proof three lemmas giving distribution theory which could be used to this end. These lemmas may be of some interest in themselves, and could possibly be shown to yield procedures in some way superior to the method of iteration.

Let $X_1, \ldots, X_k$ be independent, $X_i \sim \text{Geom}(p_i)$, and define

$$W_0 = \min(X_1, \ldots, X_k), \quad W_i = X_{i+1} - X_i, \quad i = 1, \ldots, k-1. \tag{5.15}$$

Lemma 5.4.

$W_0 \sim \text{Geom}(p_1 \cdots p_k)$ and W_0 is independent of $(W_1, \ldots, W_{k-1})$.

Lemma 5.5.

$\{W_0 = w_0\}$ is independent of $\{X_1 \leq X_2 \leq \cdots \leq X_k\}$.

Lemma 5.6.

Given that $X_1 \leq X_2 \leq \cdots \leq X_k$, $W_0, W_1, \ldots, W_{k-1}$ are mutually independent and $W_i \sim \text{Geom}(p_{i+1} \cdots p_k)$ for $i = 0, 1, \ldots, k-1$.

observations. Utilizing observations with $i = 1, 2$, we may construct by the method of Section 5.1, values $U_1, \ldots, U_r$ which are independent $\mathrm{Geom}(p_1p_2)$ variates. Next we combine $U_1, \ldots, U_r$ with $X_{31}, \ldots, X_{3r}$ to obtain $\mathrm{Geom}(p_1p_2p_3)$ variates, and so on.

Other procedures can be described which involve choosing the minimum of more than two geometric variates. It does not seem worthwhile to spell out details here, but we will state without proof three lemmas giving distribution theory which could be used to this end. These lemmas may be of some interest in themselves, and could possibly be shown to yield procedures in some way superior to the method of iteration.

Let $X_1, \ldots, X_k$ be independent, $X_i \sim \mathrm{Geom}(p_i)$, and define

$$(5.15)\qquad W_0 = \min(X_1, \ldots, X_k), \quad W_i = X_{i+1} - X_i, \quad i = 1, \ldots, k-1.$$

<u>Lemma 5.4.</u>

$W_0 \sim \mathrm{Geom}(p_1 \ldots p_k)$ and W_0 is independent of $(W_1, \ldots, W_{k-1})$.

<u>Lemma 5.5.</u>

$\{W_0 = w_0\}$ is independent of $\{X_1 \leq X_2 \leq \ldots \leq X_k\}$.

<u>Lemma 5.6.</u>

Given that $X_1 \leq X_2 \leq \ldots \leq X_k$, $W_0, W_1, \ldots, W_{k-1}$ are mutually independent and $W_i \sim \mathrm{Geom}(p_{i+1} \ldots p_k)$ for $i = 0, 1, \ldots, k-1$.

REFERENCES

[1] Bartko, John, J., "Approximating the Negative Binomial," Technometrics, 8 (1966), 345-50.

[2] Buehler, Robert J., "Confidence Limits for the Product of Two Binomial Parameters," Journal of the American Statistical Association, 52 (1957), 482-93.

[3] Feller, William, An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd ed., New York: John Wiley and Sons, Inc., 1968.

[4] Ferguson, T. S., "A Characterization of the Geometric Distribution," American Mathematical Monthly, 72 (1965), 256-60.

[5] Harris, Bernard, "Hypothesis Testing and Confidence Intervals for Products and Quotients of Poisson Parameters with Applications to Reliability," Journal of the American Statistical Association, 66 (1971), 609-13.

[6] Hwang, Dar-Shong, "Interval Estimation of Functions of Bernoulli Parameters with Reliability and Biomedical Applications," University of Minnesota PhD Thesis and School of Statistics Technical Report No. 152, Minneapolis, 1971.

[7] Johnson, Norman L. and Kotz, Samuel, Discrete Distributions, Boston: Houghton Mifflin Co., 1969.

[8] Lehmann, E. L., Testing Statistical Hypotheses, New York: John Wiley and Sons, Inc., 1959.

[9] Lehmann, E. L. and Scheffé, Henry, "Completeness, Similar Regions, and Unbiased Estimation, Part II," Sankhyā, 15 (1955), 219-36.

REFERENCES

[1] Bartko, John J., "Approximating the Negative Binomial," *Technometrics*, 2 (1960), 345-50.

[2] Buehler, Robert J., "Confidence Limits for the Product of Two Binomial Parameters," *Journal of the American Statistical Association*, 52 (1957), 482-93.

[3] Feller, William, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 2nd ed., New York: John Wiley and Sons, Inc., 1957.

[4] Ferguson, T. S., "A Characterization of the Geometric Distribution," *American Mathematical Monthly*, 72 (1965), 256-60.

[5] Harris, Bernard, "Hypothesis Testing and Confidence Intervals for Products and Quotients of Poisson Parameters with Applications to Reliability," *Journal of the American Statistical Association*, 66 (1971), 609-13.

[6] Hwang, Dar-Shong, "Interval Estimation of Functions of Bernoulli Parameters with Reliability and Biomedical Applications," University of Minnesota PhD Thesis and School of Statistics *Technical Report No. 152*, Minneapolis, 1971.

[7] Johnson, Norman L. and Kotz, Samuel, *Discrete Distributions*, Boston: Houghton Mifflin Co., 1969.

[8] Lehmann, E. L., *Testing Statistical Hypotheses*, New York: John Wiley and Sons, Inc., 1959.

[9] Lehmann, E. L. and Scheffé, Henry, "Completeness, Similar Regions, and Unbiased Estimation, Part II," *Sankhyā*, 15 (1955), 219-36.

[10] Lentner, M. M. and Buehler, R. J., "Some Inferences about Gamma Parameters with an Application to a Reliability Problem," Journal of the American Statistical Association, 58 (1963), 670-7.

[11] Lieberman, Gerald J. and Ross, Sheldon M., "Confidence Intervals for Independent Exponential Series Systems," Technical Report No. 130, Department of Operations Research and Department of Statistics, Stanford University, Stanford, California, 1970. (To appear in Journal of the American Statistical Association.)

[12] Madansky, Albert, "Approximate Confidence Limits for the Reliability of Series and Parallel Systems," Technometrics, 7 (1965), 495-503.

[13] Mann, Nancy R., "Computer-Aided Selection of Prior Distributions for Generating Monte Carlo Confidence Bounds on System Reliability," Naval Research Logistics Quarterly, 17 (1970), 41-54.

[14] Myhre, Janet and Saunders, Sam C., "On Confidence Limits for the Reliability of Systems," The Annals of Mathematical Statistics, 39 (1968), 1463-72.

[15] Patil, G. P., "On the Evaluation of the Negative Binomial Distribution with Examples," Technometrics, 2 (1960), 501-5.

[16] Pearson, E. S. and Hartley, H. O., Biometrika Tables for Statisticians, Volume I, Cambridge University Press, 1956.

[17] Pearson, Karl, Tables of the Incomplete Beta-Function, Cambridge University Press, 1934.

[18] Rosenblatt, Joan Raup, "Confidence Limits for the Reliability of Complex Systems," Statistical Theory of Reliability, 115-37, Marvin Zelen, Ed., University of Wisconsin Press, Madison, 1963.

[10] Lentner, M. M. and Buehler, R. J., "Some Inferences about Gamma Parameters with an Application to a Reliability Problem," Journal of the American Statistical Association, 58 (1963), 670-7.

[11] Lieberman, Gerald J. and Ross, Sheldon M., "Confidence Intervals for Independent Exponential Series Systems," Technical Report No. 120, Department of Operations Research and Department of Statistics, Stanford University, Stanford, California, 1970. (To appear in Journal of the American Statistical Association.)

[12] Madansky, Albert, "Approximate Confidence Limits for the Reliability of Series and Parallel Systems," Technometrics, 7 (1965), 495-503.

[13] Mann, Nancy R., "Computer-Aided Selection of Prior Distributions for Generating Monte Carlo Confidence Bounds on System Reliability," Naval Research Logistics Quarterly, 17 (1970), 41-54.

[14] Myhre, Janet and Saunders, Sam C., "On Confidence Limits for the Reliability of Systems," The Annals of Mathematical Statistics, 39 (1968), 1463-72.

[15] Patil, G. P., "On the Evaluation of the Negative Binomial Distribution with Examples," Technometrics, 2 (1960), 501-5.

[16] Pearson, Karl, and Hartley, H. O., Biometrika Tables for Statisticians, Volume I, Cambridge University Press, 1956.

[17] Pearson, Karl, Tables of the Incomplete Beta-Function, Cambridge University Press, 1934.

[18] Rosenblatt, Joan Raup, "Confidence Limits for the Reliability of Complex Systems," Statistical Theory of Reliability, 115-37. Marvin Zelen, Ed.; University of Wisconsin Press, Madison, 1963.

[19] Stevens, W. L., "Fiducial Limits of the Parameter of a Discontinuous Distribution," *Biometrika*, 37 (1950), 117-29.

[20] Whittaker, E. T. and Watson, G. N., *A Course of Modern Analysis*, 2nd ed. Cambridge, England: Cambridge University Press, 1915.